A REPORTER AT LARGE

SAM ALTMAN MAY CONTROL OUR FUTURE—CAN HE BE TRUSTED?

New interviews and closely guarded documents shed light on the persistent doubts about the head of OpenAI.

By Ronan Farrow and Andrew Marantz

April 6, 2026

Altman promised to be a safe steward for A.I. But some of his colleagues believed that he was not trustworthy enough to, as one put it, "have his finger on the button." Visual by David Szauder; Generated using A.I.

Save this story

In the fall of 2023, Ilya Sutskever, OpenAI's chief scientist, sent secret memos to three fellow-members of the organization's board of directors. For weeks, they'd been having furtive discussions about whether Sam Altman, OpenAI's C.E.O., and Greg Brockman, his second-in-command, were fit to run the company. Sutskever had once counted both men as friends. In 2019, he'd officiated Brockman's wedding, in a ceremony at OpenAI's offices that included a ring bearer in the form of a robotic hand. But as he grew convinced that the company was nearing its long-term goal—creating an artificial intelligence that could rival or surpass the cognitive capabilities of human beings—his doubts about Altman increased. As Sutskever put it to another board member at the time, "I don't think Sam is the guy who should have his finger on the button."

At the behest of his fellow board members, Sutskever worked with like-minded colleagues to compile some seventy pages of Slack messages and H.R. documents, accompanied by explanatory text. The material included images taken with a cellphone, apparently to avoid detection on company devices. He sent the final memos to the other board members as

disappearing messages, to insure that no one else would ever see them. "He was terrified," a board member who received them recalled. The memos, which we reviewed, have not previously been disclosed in full. They allege that Altman misrepresented facts to executives and board members, and deceived them about internal safety protocols. One of the memos, about Altman, begins with a list headed "Sam exhibits a consistent pattern of . . ." The first item is "Lying."

Many technology companies issue vague proclamations about improving the world, then go about maximizing revenue. But the founding premise of OpenAI was that it would have to be different. The founders, who included Altman, Sutskever, Brockman, and Elon Musk, asserted that artificial intelligence could be the most powerful, and potentially dangerous, invention in human history, and that perhaps, given the existential risk, an unusual corporate structure would be required. The firm was established as a nonprofit, whose board had a duty to prioritize the safety of humanity over the company's success, or even its survival. The C.E.O. had to be a person of uncommon integrity. According to Sutskever, "any person working to build this civilization-altering technology bears a heavy burden and is taking on unprecedented responsibility." But "the people who end up in these kinds of positions are often a certain kind of person, someone who is interested in power, a politician, someone who likes it." In one of the memos, he seemed concerned with entrusting the technology to someone who "just tells people what they want to hear." If OpenAI's C.E.O. turned out not to be reliable, the board, which had six members, was empowered to fire him. Some members, including Helen Toner, an A.I.-policy expert, and Tasha McCauley, an entrepreneur, received the memos as a confirmation of what they had already come to believe: Altman's role entrusted him with the

future of humanity, but he could not be trusted.

Altman was in Las Vegas, attending a Formula 1 race, when Sutskever invited him to a video call with the board, then read a brief statement explaining that he was no longer an employee of OpenAI. The board, following legal advice, released a public message saying only that Altman had been removed because he "was not consistently candid in his communications." Many of OpenAI's investors and executives were shocked. Microsoft, which had invested some thirteen billion dollars in OpenAI, learned of the plan to fire Altman just moments before it happened. "I was very stunned," Satya Nadella, Microsoft's C.E.O., later said. "I couldn't get anything out of anybody." He spoke with the LinkedIn co-founder Reid Hoffman, an OpenAI investor and a Microsoft board member, who began calling around to determine whether Altman had committed a clear offense. "I didn't know what the fuck was going on," Hoffman told us. "We were looking for embezzlement, or sexual harassment, and I just found nothing."

Get our Classics newsletter to discover timeless gems from *The New Yorker* archive.

SIGN UP

By signing up, you agree to our user agreement (including class action waiver and arbitration provisions), and acknowledge our privacy policy.

Other business partners were similarly blindsided. When Altman called the investor Ron Conway to say that he'd been fired, Conway held up his phone to Representative Nancy Pelosi, with whom he was having lunch. "You better get out of here really quick," she told Conway. OpenAI was on the verge of closing a large investment from Thrive, a venture-capital firm founded by Josh Kushner, Jared Kushner's brother, whom Altman had known for years. The deal would value OpenAI at eighty-six billion dollars and allow many employees to cash out millions in equity. Kushner emerged from a meeting with Rick Rubin, the music producer, to a missed call from Altman. "We just immediately went to war," Kushner later said.

The day that Altman was fired, he flew back to his twenty-seven-million-dollar mansion in San Francisco, which has panoramic views of the bay and once featured a cantilevered infinity pool, and set up what he called a "sort of government-in-exile." Conway, the Airbnb co-founder Brian Chesky, and the famously aggressive crisis-communications manager Chris Lehane joined, sometimes for hours a day, by video and phone. Some members of Altman's executive team camped out in the hallways of the house. Lawyers set up in a home office next to his bedroom. During bouts of insomnia, Altman would wander by them in his pajamas. When we spoke with Altman recently, he described the aftermath of his firing as "just this weird fugue."

With the board silent, Altman's advisers built a public case for his return. Lehane has insisted that the firing was a coup orchestrated by rogue "effective altruists"—adherents of a belief system that focusses on

maximizing the well-being of humanity, who had come to see A.I. as an existential threat. (Hoffman told Nadella that the firing might be due to "effective-altruism craziness.") Lehane—whose reported motto, after Mike Tyson, is "Everyone has a game plan until you punch them in the mouth"—urged Altman to wage an aggressive social-media campaign. Chesky stayed in contact with the tech journalist Kara Swisher, relaying criticism of the board.

Cartoon by Glen Baxter

Open cartoon gallery

Altman interrupted his "war room" at six o'clock each evening with a round of Negronis. "You need to chill," he recalls saying. "Whatever's gonna happen is gonna happen." But, he added, his phone records show that he was on calls for more than twelve hours a day. At one point, Altman conveyed to Mira Murati, who had given Sutskever material for his memos and was serving as the interim C.E.O. of OpenAI in that period, that his allies were "going all out" and "finding bad things" to damage her reputation, as well as those of others who had moved against him, according to someone with knowledge of the conversation. (Altman does not recall the exchange.)

Within hours of the firing, Thrive had put its planned investment on hold

and suggested that the deal would be consummated—and employees would thus receive payouts—only if Altman returned. Texts from this period show Altman coördinating closely with Nadella. ("how about: satya and my top priority remains to save openai," Altman suggested, as the two worked on a statement. Nadella proposed an alternative: "to ensure OpenAI continues to thrive.") Microsoft soon announced that it would create a competing initiative for Altman and any employees who left OpenAI. A public letter demanding his return circulated at the organization. Some people who hesitated to sign it received imploring calls and messages from colleagues. A majority of OpenAI employees ultimately threatened to leave with Altman.

The board was backed into a corner. "Control Z, that's one option," Toner said—undo the firing. "Or the other option is the company falls apart." Even Murati eventually signed the letter. Altman's allies worked to win over Sutskever. Brockman's wife, Anna, approached him at the office and pleaded with him to reconsider. "You're a good person—you can fix this," she said. Sutskever later explained, in a court deposition, "I felt that if we were to go down the path where Sam would not return, then OpenAI would be destroyed." One night, Altman took an Ambien, only to be awakened by his husband, an Australian coder named Oliver Mulherin, who told him that Sutskever was wavering, and that people were telling Altman to speak with the board. "I woke up in this, like, crazy Ambien haze, and I was so disoriented," Altman told us. "I was, like, I cannot talk to the board right now."

In a series of increasingly tense calls, Altman demanded the resignations of board members who had moved to fire him. "I have to pick up the pieces of their mess while I'm in this crazy cloud of suspicion?" Altman

recalled initially thinking, about his return. “I was just, like, Absolutely fucking not.” Eventually, Sutskever, Toner, and McCauley lost their board seats. Adam D’Angelo, a founder of Quora, was the sole original member who remained. As a condition of their exit, the departing members demanded that the allegations against Altman—including that he pitted executives against one another and concealed his financial entanglements—be investigated. They also pressed for a new board that could oversee the outside inquiry with independence. But the two new members, the former Harvard president Lawrence Summers and the former Facebook C.T.O. Bret Taylor, were selected after close conversations with Altman. “would you do this,” Altman texted Nadella. “bret, larry summers, adam as the board and me as ceo and then bret handles the investigation.” (McCauley later testified in a deposition that when Taylor was previously considered for a board seat she’d had concerns about his deference to Altman.)

Less than five days after his firing, Altman was reinstated. Employees now call this moment “the Blip,” after an incident in the Marvel films in which characters disappear from existence and then return, unchanged, to a world profoundly altered by their absence. But the debate over Altman’s trustworthiness has moved beyond OpenAI’s boardroom. The colleagues who facilitated his ouster accuse him of a degree of deception that is untenable for any executive and dangerous for a leader of such a transformative technology. “We need institutions worthy of the power they wield,” Murati told us. “The board sought feedback, and I shared what I was seeing. Everything I shared was accurate, and I stand behind

all of it." Altman's allies, on the other hand, have long dismissed the accusations. After the firing, Conway texted Chesky and Lehane demanding a public-relations offensive. "This is REPUTATIONAL TO SAM," he wrote. He told the Washington *Post* that Altman had been "mistreated by a rogue board of directors."

OpenAI has since become one of the most valuable companies in the world. It is reportedly preparing for an initial public offering at a potential valuation of a trillion dollars. Altman is driving the construction of a staggering amount of A.I. infrastructure, some of it concentrated within foreign autocracies. OpenAI is securing sweeping government contracts, setting standards for how A.I. is used in immigration enforcement, domestic surveillance, and autonomous weaponry in war zones.

Altman has promoted OpenAI's growth by touting a vision in which, he wrote in a 2024 blog post, "astounding triumphs—fixing the climate, establishing a space colony, and the discovery of all of physics—will eventually become commonplace." His rhetoric has helped sustain one of the fastest cash burns of any startup in history, relying on partners that have borrowed vast sums. The U.S. economy is increasingly dependent on a few highly leveraged A.I. companies, and many experts—at times including Altman—have warned that the industry is in a bubble. "Someone is going to lose a phenomenal amount of money," he told reporters last year. If the bubble pops, economic catastrophe may follow. If his most bullish projections prove correct, he may become one of the wealthiest and most powerful people on the planet.

In a tense call after Altman's firing, the board pressed him to acknowledge a pattern of deception. "This is just so fucked up," he said repeatedly, according to people on the call. "I can't change my personality." Altman says that he doesn't recall the exchange. "It's possible I meant something like 'I do try to be a unifying force,' " he told us, adding that this trait had enabled him to lead an immensely successful company. He attributed the criticism to a tendency, especially early in his career, "to be too much of a conflict avoider." But a board member offered a different interpretation of his statement: "What it meant was 'I have this trait where I lie to people, and I'm not going to stop.' " Were the colleagues who fired Altman motivated by alarmism and personal animus, or were they right that he couldn't be trusted?

One morning this winter, we met Altman at OpenAI's headquarters, in San Francisco, for one of more than a dozen conversations with him for this story. The company had recently moved into a pair of eleven-story glass towers, one of which had been occupied by Uber, another tech behemoth, whose co-founder and C.E.O., Travis Kalanick, seemed like an unstoppable prodigy—until he resigned, in 2017, under pressure from investors, who cited concerns about his ethics. (Kalanick now runs a robotics startup; in his free time, he said recently, he uses OpenAI's ChatGPT "to get to the edge of what's known in quantum physics.")

An employee gave us a tour of the office. In an airy space full of communal tables, there was an animated digital painting of the computer scientist Alan Turing; its eyes tracked us as we passed. The installation is a winking

reference to the Turing test, the 1950 thought experiment about whether a machine can credibly imitate a person. (In a 2025 study, ChatGPT passed the test more reliably than actual humans did.) Typically, you can interact with the painting. But the sound had been disabled, our guide told us, because it wouldn't stop eavesdropping on employees and then butting into their conversations. Elsewhere in the office, plaques, brochures, and merchandise displayed the words "Feel the AGI." The phrase was originally associated with Sutskever, who used it to caution his colleagues about the risks of artificial general intelligence—the threshold at which machines match human cognitive capacities. After the Blip, it became a cheerful slogan hailing a superabundant future.

We met Altman in a generic-looking conference room on the eighth floor. "People used to tell me about decision fatigue, and I didn't get it," Altman told us. "Now I wear a gray sweater and jeans every day, and even picking which gray sweater out of my closet—I'm, like, I wish I didn't have to think about that." Altman has a youthful appearance—he is slender, with wide-set blue eyes and tousled hair—but he is now forty, and he and Mulherin have a one-year-old son, delivered by a surrogate. "I'm sure, like, being President of the United States would be a much more stressful job, but of all the jobs that I think I could reasonably do, this is the most stressful one I can imagine," he said, making eye contact with one of us, then with the other. "The way that I've explained this to my friends is: 'This was the most fun job in the world until the day we launched ChatGPT.' We were making these massive scientific discoveries—I think we did the most important piece of scientific discovery in, I don't know, many decades." He cast his eyes down. "And then, since the launch of ChatGPT, the decisions have gotten very difficult."

Altman grew up in Clayton, Missouri, an affluent suburb of St. Louis, as the eldest of four siblings. His mother, Connie Gibstine, is a dermatologist; his father, Jerry Altman, was a real-estate broker and a housing activist. Altman attended a Reform synagogue and a private preparatory school that he has described as "not the kind of place where you would really stand up and talk about being gay." In general, though, the family's wealthy suburban circles were relatively liberal. When Altman was sixteen or seventeen, he said, he was out late in a predominantly gay neighborhood in St. Louis and was subjected to a brutal physical attack and homophobic slurs. Altman did not report the incident, and he was reluctant to give us more details on the record, saying that a fuller telling would "make me look like I'm manipulative or playing for sympathy." He dismissed the idea that this event, and his sexuality broadly, was significant to his identity. But, he said, "probably that has, like, some deep-seated psychological thing—that I think I'm over but I'm not—about not wanting more conflict."

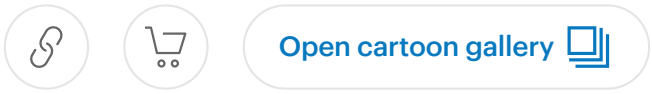
Altman's attitude in childhood, his brother told *The New Yorker*, in 2016, was "I have to win, and I'm in charge of everything." He went to Stanford, where he attended regular off-campus poker games. "I think I learned more about life and business from that than I learned in college," he later said.

All Stanford students are ambitious, but many of the most enterprising among them drop out. The summer after his sophomore year, Altman went to Massachusetts to join the inaugural batch of entrepreneurs at Y

Combinator, a "startup incubator" co-founded by the renowned software engineer Paul Graham. Each entrant joined Y.C. with an idea for a startup. (Altman's batch mates included founders of Reddit and Twitch.) Altman's project, eventually called Loopt, was a proto social network that used the locations of people's flip phones to tell their friends where they were. The company reflected his drive, and a tendency to interpret ambiguous situations to his advantage. Federal rules required that phone carriers be able to track the locations of phones for emergency services; Altman struck deals with carriers to tap these capabilities for the company's use.

"These numbers indicate that somebody here has the soul of a poet."

Cartoon by Emily Flake



Most of Altman's employees at Loopt liked him, but some said that they were struck by his tendency to exaggerate, even about trivial things. One recalled Altman bragging widely that he was a champion Ping-Pong player—"like, Missouri high-school Ping-Pong champ"—and then proving to be one of the worst players in the office. (Altman says that he was probably joking.) As Mark Jacobstein, an older Loopt employee who was asked by investors to act as Altman's "babysitter," later told Keach Hagey, for "The Optimist," a biography of Altman, "There's a blurring between 'I think I can maybe accomplish this thing' and 'I have already

accomplished this thing' that in its most toxic form leads to Theranos," Elizabeth Holmes's fraudulent startup.

Groups of senior employees, concerned with Altman's leadership and lack of transparency, asked Loopt's board on two occasions to fire him as C.E.O., according to Hagey. But Altman inspired fierce loyalty, too. A former employee was told that a board member responded, "This is Sam's company, get back to fucking work." (A board member denied that the attempts to remove Altman as C.E.O. were serious.) Loopt struggled to gain users, and in 2012 it was acquired by a fintech company. The acquisition had been arranged, according to a person familiar with the deal, largely to help Altman save face. Still, by the time Graham retired from Y.C., in 2014, he had recruited Altman to be his successor as president. "I asked Sam in our kitchen," Graham told *The New Yorker*. "And he smiled, like, *it worked*. I had never seen an uncontrolled smile from Sam. It was like when you throw a ball of paper into the wastebasket across the room—that smile."

Altman's new role made him, at twenty-eight, a kingmaker. His job was to select the hungriest and most promising entrepreneurs, connect them with the best coders and investors, and help them develop their startups into industry-defining monopolies (while Y.C. took a six- or seven-per-cent cut). Altman oversaw a period of aggressive expansion, growing Y.C.'s roster of startups from dozens to hundreds. But several Silicon Valley investors came to believe that his loyalties were divided. An investor told us that Altman was known to "make personal investments, selectively, into

the best companies, blocking outside investors." (Altman denies blocking anyone.) Altman had worked as a "scout" for the investment fund Sequoia Capital, as part of a program that involved investing in early-stage startups and taking a small cut of any profits. When Altman made an angel investment in Stripe, a financial-services startup, he insisted on a bigger portion, galling Sequoia's partners, a person familiar with the deal said. The person added, "It's a policy of 'Sam first.' " Altman is an investor in, by his own estimate, some four hundred other companies. (Altman denies this characterization of the Stripe deal. Around 2010, he made an initial investment of fifteen thousand dollars in Stripe, a two-per-cent share. The company is now valued at more than a hundred and fifty billion dollars.)

By 2018, several Y.C. partners were so frustrated with Altman's behavior that they approached Graham to complain. Graham and Jessica Livingston, his wife and a Y.C. founder, apparently had a frank conversation with Altman. Afterward, Graham started telling people that although Altman had agreed to leave the company, he was resisting in practice. Altman told some Y.C. partners that he would resign as president but become chairman instead. In May, 2019, a blog post announcing that Y.C. had a new president came with an asterisk: "Sam is transitioning to Chairman of YC." A few months later, the post was edited to read "Sam Altman stepped away from any formal position at YC"; after that, the phrase was removed entirely. Nevertheless, as recently as 2021, a Securities and Exchange Commission filing listed Altman as the chairman of Y Combinator. (Altman says that he wasn't aware of this until much later.)

Altman has maintained over the years, both in public and in recent depositions, that he was never fired from Y.C., and he told us that he did not resist leaving. Graham has tweeted that "we didn't want him to leave, just to choose" between Y.C. and OpenAI. In a statement, Graham told us, "We didn't have the legal power to fire anyone. All we could do was apply moral pressure." In private, though, he has been unambiguous that Altman was removed because of Y.C. partners' mistrust. This account of Altman's time at Y Combinator is based on discussions with several Y.C. founders and partners, in addition to contemporaneous materials, all of which indicate that the parting was not entirely mutual. On one occasion, Graham told Y.C. colleagues that, prior to his removal, "Sam had been lying to us all the time."

In May, 2015, Altman e-mailed Elon Musk, then the hundredth-richest person in the world. Like many prominent Silicon Valley entrepreneurs, Musk was preoccupied by an array of threats that he considered existentially urgent but which would have struck most people as far-fetched hypotheticals. "We need to be super careful with AI," he tweeted. "Potentially more dangerous than nukes."

Altman had generally been a techno-optimist, but his rhetoric about A.I. soon turned apocalyptic. In public, and in his private correspondence with Musk and others, he warned that the technology should not be dominated by a profit-seeking mega-corporation. "Been thinking a lot about whether it's possible to stop humanity from developing AI," he wrote to Musk. "If it's going to happen anyway, it seems like it would be good for someone

other than Google to do it first." Picking up on the analogy to nuclear weapons, he proposed a "Manhattan Project for AI." He outlined the overarching principles that such an organization would have—"safety should be a first-class requirement"; "obviously we'd comply with/aggressively support all regulation"—and he and Musk settled on a name: OpenAI.

Unlike the original Manhattan Project, a government initiative that led to the creation of the atom bomb, OpenAI would be privately funded, at least at first. Altman predicted that an artificial superintelligence—a theoretical threshold beyond even A.G.I., at which machines would fully eclipse the capabilities of the human mind—would eventually create enough economic benefits to "capture the light cone of all future value in the universe." But he also warned of existential danger. At some point, the national-security implications could grow so dire that the U.S. government would have to take control of OpenAI, perhaps by nationalizing it and moving its operations to a secure bunker in the desert. By late 2015, Musk was persuaded. "We should say that we are starting with a $1B funding commitment," he wrote. "I will cover whatever anyone else doesn't provide."

Altman housed OpenAI in Y Combinator's nonprofit arm, framing it as an internal philanthropic project. He gave OpenAI recruits Y.C. stock and moved donations through Y.C. accounts. At one point, the lab was supported by a Y.C. fund in which he held a personal stake. (Altman later described this stake as insignificant. He told us that the Y.C. stock he gave

to recruits was his own.)

The Manhattan Project analogy applied to employee recruitment, too. Like nuclear-fission research, machine learning was a small scientific field with epochal implications which was dominated by a cadre of eccentric geniuses. Musk and Altman, along with Brockman, who joined from Stripe, were convinced that there were only a few computer scientists alive capable of making the required breakthroughs. Google had a huge cash advantage and a multiyear head start. “We are outmanned and outgunned by a ridiculous margin,” Musk later wrote in an e-mail. But “if we are able to attract the most talented people over time and our direction is correctly aligned, then OpenAI will prevail.”

A top recruiting target was Sutskever, an intense and introverted researcher who was often called the most gifted A.I. scientist of his generation. Sutskever, who was born in the Soviet Union in 1986, has a receding hairline, dark eyes, and a habit of pausing, unblinking, while choosing his words. Another target was Dario Amodei, a biophysicist and a font of frenetic energy who has a tendency to nervously twist his black hair, and responds to one-line e-mails with multi-paragraph essays. Both had lucrative jobs elsewhere, but Altman lavished them with attention. He later joked, “I stalked Ilya.”

Musk was the bigger name, but Altman had the smoother touch. He e-mailed Amodei, and they set up a one-on-one dinner at an Indian restaurant. (Altman: “fuck my uber got in a crash! running about 10 late.” Amodei: “Wow, hope you’re ok.”) Like many A.I. researchers, Amodei

believed that the technology should be built only if it was shown to be "aligned" with human values, meaning that it would act in accordance with what people wanted without making a potentially fatal error—say, following an instruction to clean up the environment by eliminating its greatest polluter, the human race. Altman was reassuring, mirroring these safety concerns.

Amodei, who later joined the company, took detailed notes on Altman and Brockman's behavior for years, under the heading "My Experience with OpenAI" (subheading: "Private: Do Not Share"). A collection of more than two hundred pages of documents related to Amodei, including those notes and internal e-mails and memos, has been circulated by colleagues in Silicon Valley but never before disclosed publicly. In his notes, Amodei wrote that Altman's goal was to build "an AI lab that would be focused on safety ('maybe not right away, but as soon as it can be')."

In December, 2015, hours before OpenAI was publicly announced, Altman e-mailed Musk about a rumor that Google was "going to give everyone in openAI massive counteroffers tomorrow to try to kill it." Musk replied, "Has Ilya come back with a solid yes?" Altman assured him that Sutskever was holding firm. Google offered Sutskever six million dollars a year, which OpenAI couldn't come close to matching. But, Altman boasted, "they unfortunately dont have 'do the right thing' on their side."

"I'm just saying, if we tear up the pillows and rip up the mattress, it might make our place look more lived in."

Cartoon by Lonnie Millsap

Open cartoon gallery

Musk provided some office space for OpenAI in a former suitcase factory in the Mission District of San Francisco. The pitch to employees, Sutskever told us, was "You're going to save the world."

If everything went right, the OpenAI founders believed, artificial intelligence could usher in a post-scarcity utopia, automating grunt work, curing cancer, and liberating people to enjoy lives of leisure and abundance. But if the technology went rogue, or fell into the wrong hands, the devastation could be total. China could use it to build a novel bioweapon or a fleet of advanced drones; an A.I. model could outmaneuver its overseers, replicating itself on secret servers so that it couldn't be turned off; in extreme cases, it might seize control of the energy grid, the stock market, or the nuclear arsenal. Not everyone believed this, to say the least, but Altman repeatedly affirmed that he did. He wrote on his blog in 2015 that superhuman machine intelligence "does not have to be the inherently evil sci-fi version to kill us all. A more probable scenario is that it simply doesn't care about us much either way, but in an effort to accomplish some other goal . . . wipes us out." OpenAI's founders vowed not to privilege speed over safety, and the organization's articles of incorporation made benefitting humanity a legally binding duty. If A.I. was going to be the most powerful technology in history, it

followed that any individual with sole control over it stood to become uniquely powerful—a scenario that the founders referred to as an "AGI dictatorship."

Altman told early recruits that OpenAI would remain a pure nonprofit, and programmers took significant pay cuts to work there. The company accepted charitable grants, including thirty million dollars from what was then called Open Philanthropy, a hub of the effective-altruism movement whose commitments included supporting the distribution of mosquito nets to the global poor.

Brockman and Sutskever managed OpenAI's daily operations, while Musk and Altman, still busy with their other jobs, stopped by around once a week. By September, 2017, though, Musk had grown impatient. During discussions about whether to reconstitute OpenAI as a for-profit company, he demanded majority control. Altman's replies varied depending on the context. His main consistent demand seems to have been that if OpenAI were reorganized under the control of a C.E.O. that job should go to him. Sutskever seemed uncomfortable with this idea. He sent Musk and Altman a long, plaintive e-mail on behalf of himself and Brockman, with the subject line "Honest Thoughts." He wrote, "The goal of OpenAI is to make the future good and to avoid an AGI dictatorship." He continued, addressing Musk, "So it is a bad idea to create a structure where you could become a dictator." He relayed similar concerns to Altman: "We don't understand why the CEO title is so important to you. Your stated reasons have changed, and it's hard to really understand what's driving it."

“Guys, I’ve had enough,” Musk replied. “Either go do something on your own or continue with OpenAI as a nonprofit”—otherwise “I’m just being a fool who is essentially providing free funding for you to create a startup.” He quit, acrimoniously, five months later. (In 2023, he founded a for-profit competitor called xAI. The following year, he sued Altman and OpenAI for fraud and breach of charitable trust, alleging that he had been “assiduously manipulated” by “Altman’s long con”—that Altman had preyed on his concerns about the dangers of A.I. in order to separate him from his money. The suit, which OpenAI has vigorously contested, is ongoing.)

After Musk’s departure, Amodei and other researchers chafed against the leadership of Brockman, whom some considered an abrasive operator, and of Sutskever, who was generally viewed as principled but disorganized. In the process of becoming C.E.O., Altman seems to have made different promises to different factions at the company. He assured some researchers that Brockman’s managerial authority would be diminished. But, unbeknownst to them, he also struck a secret handshake deal with Brockman and Sutskever: Altman would get the C.E.O. title; in exchange, he agreed to resign if the other two deemed it necessary. (He disputed this characterization, saying he took the C.E.O. role only because he was asked to. All three men confirmed that the pact existed, though Brockman said that it was informal. “He unilaterally told us that he’d step down if we ever both asked him to,” he told us. “We objected to this idea, but he said it was important to him. It was purely altruistic.”) Later, the board was alarmed to learn that its C.E.O. had essentially appointed his own shadow

board.

Internal records show that the founders had private doubts about the nonprofit structure as early as 2017. That year, after Musk tried to take control, Brockman wrote in a diary entry, "cannot say that we are committed to the non-profit . . . if three months later we're doing b-corp then it was a lie." Amodei, in one of his early notes, recalled pressing Brockman on his priorities and Brockman replying that he wanted "money and power." Brockman disputes this. His diary entries from this time suggest conflicting instincts. One reads, "Happy to not become rich on this, so long as no one else is." In another, he asks, "So what do I *really* want?" Among his answers is "Financially what will take me to $1B."

In 2017, Sutskever was in the office when he read a paper that Google researchers had just published, proposing "a new simple network architecture, the Transformer." He jumped out of his chair, ran down the hall, and told his fellow-researchers, "Stop everything you're doing. This is it." The Transformer, Sutskever saw, was an innovation that might enable OpenAI to train vastly more sophisticated models. Out of this discovery came the first generative pre-trained transformer—the seed of what would become ChatGPT.

As the technology became increasingly powerful, we learned, about a dozen of OpenAI's top engineers held a series of secret meetings to discuss whether OpenAI's founders, including Brockman and Altman, could be trusted. At one, an employee was reminded of a sketch by the

British comedy duo Mitchell and Webb, in which a Nazi soldier on the Eastern Front, in a moment of clarity, asks, "Are we the baddies?"

By 2018, Amodei had started questioning the founders' motives more openly. "Everything was a rotating set of schemes to raise money," he later wrote in his notes. "I felt like what OpenAI needed was a clear statement of what it would do, what it would not do, and how its existence would make the world better." OpenAI already had a mission statement: "To ensure that artificial general intelligence benefits all of humanity." But it wasn't clear to Amodei what this meant to the executives, if it meant anything at all. In early 2018, Amodei has said, he started drafting a charter for the company and, in weeks of conversations with Altman and Brockman, advocated for its most radical clause: if a "value-aligned, safety-conscious project" came close to building an A.G.I. before OpenAI did, the company would "stop competing with and start assisting this project." According to the "merge and assist" clause, as it was called, if, say, Google's researchers figured out how to build a safe A.G.I. first, then OpenAI could wind itself down and donate its resources to Google. By any normal corporate logic, this was an insane thing to promise. But OpenAI was not supposed to be a normal company.

That premise was tested in the spring of 2019, when OpenAI was negotiating a billion-dollar investment from Microsoft. Although Amodei, who was leading the company's safety team, had helped to pitch the deal to Bill Gates, many people on the team were anxious about it, fearing that Microsoft would insert provisions that overrode OpenAI's

ethical commitments. Amodei presented Altman with a ranked list of safety demands, placing the preservation of the merge-and-assist clause at the very top. Altman agreed to that demand, but in June, as the deal was closing, Amodei discovered that a provision granting Microsoft the power to block OpenAI from any mergers had been added. "Eighty per cent of the charter was just betrayed," Amodei recalled. He confronted Altman, who denied that the provision existed. Amodei read it aloud, pointing to the text, and ultimately forced another colleague to confirm its existence to Altman directly. (Altman doesn't remember this.) Amodei's notes describe escalating tense encounters, including one, months later, in which Altman summoned him and his sister, Daniela, who worked in safety and policy at the company, to tell them that he had it on "good authority" from a senior executive that they had been plotting a coup. Daniela, the notes continue, "lost it," and brought in that executive, who denied having said anything. As one person briefed on the exchange recalled, Altman then denied having made the claim. "I didn't even say that," he said. "You just said that," Daniela replied. (Altman said that this was not quite his recollection, and that he had accused the Amodeis only of "political behavior.") In 2020, Amodei, Daniela, and other colleagues left to found Anthropic, which is now one of OpenAI's chief rivals.

Altman continued touting OpenAI's commitment to safety, especially when potential recruits were within earshot. In late 2022, four computer scientists published a paper motivated in part by concerns about "deceptive alignment," in which sufficiently advanced models might pretend to behave well during testing and then, once deployed, pursue

their own goals. (It's one of several A.I. scenarios that sound like science fiction—but, under certain experimental conditions, it's already happening.) Weeks after the paper was published, one of its authors, a Ph.D. student at the University of California, Berkeley, got an e-mail from Altman, who said that he was increasingly worried about the threat of unaligned A.I. He added that he was thinking of committing a billion dollars to the issue, which many A.I. experts considered the most important unsolved problem in the world, potentially by endowing a prize to incentivize researchers around the world to study it. Although the graduate student had "heard vague rumors about Sam being slippery," he told us, Altman's show of commitment won him over. He took an academic leave to join OpenAI.

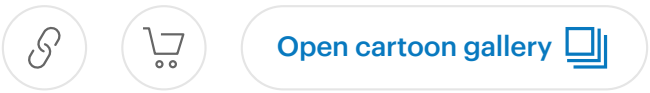
But, in the course of several meetings in the spring of 2023, Altman seemed to waver. He stopped talking about endowing a prize. Instead, he advocated for establishing an in-house "superalignment team." An official announcement, referring to the company's reserves of computing power, pledged that the team would get "20% of the compute we've secured to date"—a resource potentially worth more than a billion dollars. The effort was necessary, according to the announcement, because, if alignment remained unsolved, A.G.I. might "lead to the disempowerment of humanity or even human extinction." Jan Leike, who was appointed to lead the team with Sutskever, told us, "It was a pretty effective retention tool."

The twenty-per-cent commitment evaporated, however. Four people who

worked on or closely with the team said that the actual resources were between one and two per cent of the company's compute. Furthermore, a researcher on the team said, "most of the superalignment compute was actually on the oldest cluster with the worst chips." The researchers believed that superior hardware was being reserved for profit-generating activities. (OpenAI disputes this.) Leike complained to Murati, then the company's chief technology officer, but she told him to stop pressing the point—the commitment had never been realistic.

"She skippidy-boop-bee-bop-doo-woppity-wopped right out of my life."

Cartoon by Sofia Warren



Around this time, a former employee told us, Sutskever "was getting super safety-pilled." In the early days of OpenAI, he had considered concerns about catastrophic risk legitimate but remote. Now, as he came to believe that A.G.I. was imminent, his worries grew more acute. There was an all-hands meeting, the former employee continued, "where Ilya gets up and he's, like, Hey, everyone, there's going to be a point in the next few years where basically everyone at this company has to switch to working on safety, or else we're fucked." But the superalignment team was dissolved the following year, without completing its mission.

By then, internal messages show, executives and board members had come to believe that Altman's omissions and deceptions might have

ramifications for the safety of OpenAI's products. In a meeting in December, 2022, Altman assured board members that a variety of features in a forthcoming model, GPT-4, had been approved by a safety panel. Toner, the board member and A.I.-policy expert, requested documentation. She learned that the most controversial features—one that allowed users to "fine-tune" the model for specific tasks, and another that deployed it as a personal assistant—had not been approved. As McCauley, the board member and entrepreneur, left the meeting, an employee pulled her aside and asked if she knew about "the breach" in India. Altman, during many hours of briefing with the board, had neglected to mention that Microsoft had released an early version of ChatGPT in India without completing a required safety review. "It just was kind of completely ignored," Jacob Hilton, an OpenAI researcher at the time, said.

Although these lapses did not cause security crises, Carroll Wainwright, another researcher, said that they were part of a "continual slide toward emphasizing products over safety." After the release of GPT-4, Leike e-mailed members of the board. "OpenAI has been going off the rails on its mission," he wrote. "We are prioritizing the product and revenue above all else, followed by AI capabilities, research and scaling, with alignment and safety coming third." He continued, "Other companies like Google are learning that they should deploy faster and ignore safety problems."

McCauley, in an e-mail to her fellow-members, wrote, "I think we're definitely at a point where the board should be increasing its level of scrutiny." The board members tried to confront what they viewed as a

mounting problem, but they were outmatched. “You had a bunch of J.V. people who’ve never done anything, to be blunt,” Sue Yoon, a former board member, said. In 2023, the company was preparing to release its GPT-4 Turbo model. As Sutskever details in the memos, Altman apparently told Murati that the model didn’t need safety approval, citing the company’s general counsel, Jason Kwon. But when she asked Kwon, over Slack, he replied, “ugh . . . confused where sam got that impression.” (A representative for OpenAI, where Kwon remains an executive, said that the matter was “not a big deal.”)

Soon afterward, the board made its decision to fire Altman—and then the world watched as Altman reversed it. A version of the OpenAI charter is still on the organization’s website. But people familiar with OpenAI’s governing documents said that it has been diluted to the point of meaninglessness. Last June, on his personal blog, Altman wrote, referring to artificial superintelligence, “We are past the event horizon; the takeoff has started.” This was, according to the charter, arguably the moment when OpenAI might stop competing with other companies and start working with them. But in that post, called “The Gentle Singularity,” he adopted a new tone, replacing existential terror with ebullient optimism. “We’ll all get better stuff,” he wrote. “We will build ever-more-wonderful things for each other.” He acknowledged that the alignment problem remained unsolved, but he redefined it—rather than being a deadly threat, it was an inconvenience, like the algorithms that tempt us to waste time scrolling on Instagram.

Altman is often described, either with reverence or with suspicion, as the greatest pitchman of his generation. Steve Jobs, one of his idols, was said to project a “reality-distortion field”—an unassailable confidence

that the world would conform to his vision. But even Jobs never told his customers that if they didn't buy his brand of MP3 player everyone they loved would die. When Altman was twenty-three, in 2008, Graham, his mentor, wrote, "You could parachute him into an island full of cannibals and come back in 5 years and he'd be the king." This judgment was based not on Altman's track record, which was modest, but on his will to prevail, which Graham considered almost ungovernable. When advised not to include Y.C. alumni on a list of the world's top startup founders, Graham put Altman on it anyway. "Sam Altman can't be stopped by such flimsy rules," he wrote.

Graham meant this as a compliment. But some of Altman's closest colleagues came to have a different view of this quality. After Sutskever grew more distressed about A.I. safety, he compiled the memos about Altman and Brockman. They have since taken on a legendary status in Silicon Valley; in some circles, they are simply called the Ilya Memos. Meanwhile, Amodei was continuing to assemble notes. These and the other documents related to him chart his shift from cautious idealism to alarm. His language is more heated than Sutskever's, by turns incensed at Altman—"His words were almost certainly bullshit"—and wistful about what he says was a failure to correct OpenAI's course.

Neither collection of documents contains a smoking gun. Rather, they recount an accumulation of alleged deceptions and manipulations, each of which might, in isolation, be greeted with a shrug: Altman purportedly offers the same job to two people, tells contradictory stories about who should appear on a live stream, dissembles about safety requirements. But Sutskever concluded that this kind of behavior "does not create an environment conducive to the creation of a safe AGI." Amodei and

Sutskever were never close friends, but they reached similar conclusions. Amodei wrote, "The problem with OpenAI is Sam himself."

We have interviewed more than a hundred people with firsthand knowledge of how Altman conducts business: current and former OpenAI employees and board members; guests and staffers at Altman's various houses; his colleagues and competitors; his friends and enemies and several people who, given the mercenary culture of Silicon Valley, have been both. (OpenAI has an agreement with Condé Nast, the owner of *The New Yorker*, which allows OpenAI to display its content in search results for a limited term.)

Some people defended Altman's business acumen and dismissed his rivals, especially Sutskever and Amodei, as failed aspirants to his throne. Others portrayed them as gullible, absent-minded scientists, or as hysterical "doomers," gripped by the delusion that the software they were building would somehow come alive and kill them. Yoon, the former board member, argued that Altman was "not this Machiavellian villain" but merely, to the point of "fecklessness," able to convince himself of the shifting realities of his sales pitches. "He's too caught up in his own self-belief," she said. "So he does things that, if you live in the real world, make no sense. But he doesn't live in the real world."

Yet most of the people we spoke to shared the judgment of Sutskever and Amodei: Altman has a relentless will to power that, even among industrialists who put their names on spaceships, sets him apart. "He's unconstrained by truth," the board member told us. "He has two traits that are almost never seen in the same person. The first is a strong desire to please people, to be liked in any given interaction. The second is almost a

sociopathic lack of concern for the consequences that may come from deceiving someone."

The board member was not the only person who, unprompted, used the word "sociopathic." One of Altman's batch mates in the first Y Combinator cohort was Aaron Swartz, a brilliant but troubled coder who died by suicide in 2013 and is now remembered in many tech circles as something of a sage. Not long before his death, Swartz expressed concerns about Altman to several friends. "You need to understand that Sam can never be trusted," he told one. "He is a sociopath. He would do anything." Multiple senior executives at Microsoft said that, despite Nadella's long-standing loyalty, the company's relationship with Altman has become fraught. "He has misrepresented, distorted, renegotiated, reneged on agreements," one said. Earlier this year, OpenAI reaffirmed Microsoft as the exclusive cloud provider for its "stateless"—or memoryless—models. That day, it announced a fifty-billion-dollar deal making Amazon the exclusive reseller of its enterprise platform for A.I. agents. While reselling is permitted, Microsoft executives argue OpenAI's plan could collide with Microsoft's exclusivity. (OpenAI maintains that the Amazon deal will not violate the earlier contract; a Microsoft representative said the company is "confident that OpenAI understands and respects" its legal obligations.) The senior executive at Microsoft said, of Altman, "I think there's a small but real chance he's eventually remembered as a Bernie Madoff- or Sam Bankman-Fried-level scammer."

Altman is not a technical savant—according to many in his orbit, he lacks extensive expertise in coding or machine learning. Multiple engineers recalled him misusing or confusing basic technical terms. He built OpenAI, in large part, by harnessing other people's money and technical

talent. This doesn't make him unique. It makes him a businessman. More remarkable is his ability to convince skittish engineers, investors, and a tech-skeptical public that their priorities, even when mutually exclusive, are also his priorities. When such people have tried to hinder his next move, he has often found the words to neutralize them, at least temporarily; usually, by the time they lose patience with him, he's got what he needs. "He sets up structures that, on paper, constrain him in the future," Wainwright, the former OpenAI researcher, said. "But then, when the future comes and it comes time to be constrained, he does away with whatever the structure was."

"He's unbelievably persuasive. Like, Jedi mind tricks," a tech executive who has worked with Altman said. "He's just next level." A classic hypothetical scenario in alignment research involves a contest of wills between a human and a high-powered A.I. In such a contest, researchers usually argue, the A.I. would surely win, much the way a grandmaster will beat a child at chess. Watching Altman outmaneuver the people around him during the Blip, the executive continued, had been like watching "an A.G.I. breaking out of the box."

Cartoon by Liana Finck

Open cartoon gallery

In the days after his firing, Altman fought to avoid any outside investigation of the claims against him. He told two people that he

worried even the existence of an investigation would make him look guilty. (Altman denies this.) But, after the resigning board members made their departure conditional on there being an independent inquiry, Altman acceded to a "review" of "recent events." The two new board members insisted that they control that review, according to people involved in the negotiations. Summers, with his network of political and Wall Street connections, seemed to lend it credibility. (Last November, after the disclosure of e-mails in which Summers sought Jeffrey Epstein's advice while pursuing a romantic relationship with a young protégée, he resigned from the board.) OpenAI enlisted WilmerHale, the distinguished law firm responsible for the internal investigations of Enron and WorldCom, to conduct the review.

Six people close to the inquiry alleged that it seemed designed to limit transparency. Some of them said that the investigators initially did not contact important figures at the company. An employee reached out to Summers and Taylor to complain. "They were just interested in the narrow range of what happened during the board drama, and not the history of his integrity," the employee recalled of his interview with investigators. Others were uncomfortable sharing concerns about Altman because they felt there was not a sufficient effort to insure anonymity. "Everything pointed to the fact that they wanted to find the outcome, which is to acquit him," the employee said. (Some of the lawyers involved defended the process, saying, "It was an independent, careful, comprehensive review that followed the facts wherever they led." Taylor also said that the review was "thorough and independent.")

Corporate investigations aim to confer legitimacy. At private companies, their findings are sometimes not written down—this can be a way to limit

liability. But in cases involving public scandals there is often a greater expectation of transparency. Before Kalanick left Uber, in 2017, its board hired an outside firm, which released a thirteen-page summary to the public. Given OpenAI's 501(c)(3) status and the high-profile nature of the firing, many executives there expected to see extensive findings. In March, 2024, however, OpenAI announced that it would clear Altman but released no report. The company provided, on its website, some eight hundred words acknowledging a "breakdown in trust."

People involved in the investigation said that no report was released because none was written. Instead, the findings were limited to oral briefings, shared with Summers and Taylor. "The review did not conclude that Sam was a George Washington cherry tree of integrity," one of the people close to the inquiry said. But the investigation appears not to have centered the questions of integrity behind Altman's firing, devoting much of its focus to a hunt for clear criminality; on that basis, it concluded that he could remain as C.E.O. Shortly thereafter, Altman, who had been kicked off the board when he was fired, rejoined it. The decision not to put the report in writing was made in part on the advice of Summer's and Taylor's personal attorneys, the person close to the inquiry told us. (Summers declined to comment on the record. Taylor said that, in light of the oral briefings, there had been "no need for a formal written report.")

Many former and current OpenAI employees told us that they were shocked by the lack of disclosure. Altman said he believed that all the board members who joined in the aftermath of his reinstatement received the oral briefings. "That's an absolute, outright lie," a person with direct knowledge of the situation said. Some board members told us that ongoing questions about the integrity of the report could prompt, as one

put it, “a need for another investigation.”

The absence of a written record helped minimize the allegations. So, increasingly, did Altman’s stature in Silicon Valley. Multiple prominent investors who have worked with Altman told us that he has a reputation for freezing out investors if they back OpenAI’s competitors. “If they invest in something that he doesn’t like, they won’t get access to other things,” one of them said. Another source of Altman’s power is his vast list of investments, which at times extends to his personal life. He has financial entanglements with numerous former romantic partners: as a fund co-manager, a lead investor, or a frequent co-investor. This is hardly unusual. Many of Silicon Valley’s straight executives do the same thing with their romantic and sexual partners. (“You have to,” one prominent C.E.O. told us.) “I’ve obviously invested with some exes after the fact. And I think that’s, like, totally fine,” Altman said. But the dynamic affords an extraordinary level of control. “It creates a very, very high dependence, essentially,” a person close to Altman said. “Oftentimes, it’s a lifetime dependence.”

Even former colleagues can be affected. Murati left OpenAI in 2024 and began building her own A.I. startup. Josh Kushner, the close Altman ally, called her. He praised her leadership, then made what seemed to be a veiled threat, noting that he was “concerned about” her “reputation” and that former colleagues now viewed her as an “enemy.” (Kushner, through a representative, said that this account did not “convey full context”; Altman said that he was unaware of the call.)

At the beginning of his tenure as C.E.O., Altman had announced that OpenAI would create a “capped profit” company, which would be owned

by the nonprofit. This byzantine corporate structure apparently did not exist until Altman devised it. In the midst of the conversion, a board member named Holden Karnofsky objected to it, arguing that the nonprofit was being severely undervalued. "I can't do that in good faith," Karnofsky, who is Amodei's brother-in-law, said. According to contemporaneous notes, he voted against it. However, after an attorney for the board said that his dissent "might be a flag to investigate further" the legitimacy of the new structure, his vote was recorded as an abstention, apparently without his consent—a potential falsification of business records. (OpenAI told us that several employees recall Karnofsky abstaining, and provided the minutes from the meeting recording his vote as an abstention.)

Last October, OpenAI "recapitalized" as a for-profit entity. The firm touts its associated nonprofit, now called the OpenAI Foundation, as one of the "best resourced" in history. But it is now a twenty-six-per-cent stakeholder of the company, and its board members are also, with one exception, members of the for-profit board.

During congressional testimony, Altman was asked if he made "a lot of money." He replied, "I have no equity in OpenAI . . . I'm doing this because I love it"—a careful answer, given his indirect equity through the Y.C. fund. This is still technically true. But several people, including Altman, indicated to us that it could soon change. "Investors are, like, I need to know you're gonna stick with this when times get hard," Altman said, but added that there was no "active discussion" about it. According to a legal deposition, Brockman seems to own a stake in the company that is worth about twenty billion dollars. Altman's share would presumably be worth more. Still, he told us that he was not primarily motivated by

wealth. A former employee recalls him saying, "I don't care about money. I care more about power."

In 2023, Altman married Mulherin in a small ceremony at a home they own in Hawaii. (They'd met nine years prior, late at night in Peter Thiel's hot tub.) They have hosted a range of guests at the property, and those we spoke with reported witnessing nothing more remarkable than the standard diversions of the very wealthy: meals prepared by a private chef, boat rides at golden hour. One New Year's party was "Survivor"-themed; a photograph shows a number of shirtless, smiling men, and also Jeff Probst, the real host of "Survivor." Altman has also hosted smaller groups of friends at his properties, gatherings that have included, in at least one instance, a spirited game of strip poker. (A photograph of the event, which did not include Altman, leaves unclear who won, but at least three men clearly lost.) We spoke to many of Altman's former guests who suggested only that he is a generous host.

Nevertheless, rumors about Altman's personal life have been exploited and distorted by competitors. Ruthless business rivalries are nothing new, but the competition within the A.I. industry has become extraordinarily cutthroat. ("Shakespearean" was the word an OpenAI executive used to describe it to us, adding, "The normal rules of the game sort of don't apply anymore.") Intermediaries directly connected to, and in at least one case compensated by, Musk have circulated dozens of pages of detailed opposition research about Altman. They reflect extensive surveillance, documenting shell companies associated with him, the personal contact information of close associates, and even interviews about a purported sex worker, conducted at gay bars. One of the Musk intermediaries claimed that Altman's flights and the parties he attended were being tracked.

Altman told us, “I don’t think anyone has had more private investigators hired against them.”

Extreme claims have circulated. The right-wing broadcaster Tucker Carlson suggested, without any apparent proof, that Altman was involved in the death of a whistle-blower. This claim and others have been amplified by rivals. Altman’s sister, Annie, claimed in a lawsuit, and in interviews with us, that he sexually abused her for years, beginning when she was three and he was twelve. (We could not substantiate Annie’s account, which Altman has denied and his brothers and mother have called “utterly untrue” and a source of “immense pain to our entire family.” In interviews that the journalist Karen Hao conducted for her book, “Empire of AI,” Annie suggested that memories of abuse were recovered during flashbacks in adulthood.)

Multiple people working within rival companies and investment firms insinuated to us that Altman sexually pursues minors—a narrative persistent in Silicon Valley which appears to be untrue. We spent months looking into the matter, conducting dozens of interviews, and could find no evidence to support it. “This is disgusting behavior from a competitor that I assume is part of an attempt at tainting the jury in our upcoming cases,” Altman told us. “As ridiculous as this is to have to say, any claims about me having sex with a minor, hiring sex workers, or being involved in a murder are completely untrue.” He added that he was “sort of grateful” that we had spent months “so aggressively trying to look into this.”

“My apartment is full of smells that I personally am in no way responsible

for.”

Cartoon by William Haefeli

Altman has acknowledged dating younger men of legal age. We spoke to several of his partners, who told us that they did not find this problematic. Yet the opposition dossiers from Musk intermediaries spin it as a line of attack. (The dossiers include salacious and unsubstantiated references to a “Twink Army” and “Sugar Daddy’s Sexual Habits.”) “I think there’s a lot of homophobia that gets pushed,” Altman said. Swisher, the tech journalist, agreed. “All these rich guys do wild stuff, wilder than anything I’ve been told about Sam,” she told us. “But he’s a gay guy in San Francisco,” she added, “so that gets weaponized.”

For a decade, social-media executives promised that they could change the world with little or no downside. They dismissed the lawmakers who wanted to slow them down as mere Luddites, eventually earning bipartisan derision. Altman, by contrast, came across as refreshingly conscientious. Rather than warding off regulation, he practically begged for it. Testifying before the Senate Judiciary Committee in 2023, he proposed a new federal agency to oversee advanced A.I. models. “If this technology goes wrong, it can go quite wrong,” he said. Senator John Kennedy, of Louisiana, known for his cantankerous exchanges with tech C.E.O.s, seemed charmed, resting his face on his hand and suggesting that perhaps Altman should enforce the rules himself.

But, as Altman publicly welcomed regulation, he quietly lobbied against it.

In 2022 and 2023, according to *Time*, OpenAI successfully pressed to dilute a European Union effort that would have subjected large A.I. companies to more oversight. In 2024, a bill was introduced in the California state legislature mandating safety testing for A.I. models. Its provisions included measures resembling the ones that Altman had advocated for in his congressional testimony. OpenAI publicly opposed the bill but in private began issuing threats. "I would say that, over the course of the year, we saw increasingly cunning, deceptive behavior from OpenAI," a legislative aide told us.

Conway, the investor, lobbied state political leaders, including Nancy Pelosi and Gavin Newsom, to kill the bill. In the end, it passed the legislature with bipartisan support, but Newsom vetoed it. This year, congressional candidates who favor A.I. regulations have faced opponents funded by Leading the Future, a new "pro-A.I." super PAC devoted to scuttling such restrictions. OpenAI's official stance is that it will not contribute to such super PACs. "This issue transcends partisan politics," Lehane recently told CNN. And yet one of the major donors to Leading the Future is Greg Brockman, who has committed fifty million dollars. (This year, Brockman and his wife donated twenty-five million dollars to MAGA Inc., a pro-Trump super PAC.)

OpenAI's campaign has extended beyond traditional lobbying. Last year, a successor bill was introduced in the California Senate. One night, Nathan Calvin, a twenty-nine-year-old lawyer who worked at the nonprofit Encode and had helped craft the bill, was at home having dinner with his wife when a process server arrived to deliver a subpoena from OpenAI. The company claimed to be hunting for evidence that Musk was covertly funding its critics. But it demanded all of Calvin's private communications

about the bill in the state Senate. "They could have asked us, 'Have you ever talked to or been given money by Elon Musk?'—which we haven't," Calvin told us. Other supporters of the bill, and some critics of OpenAI's for-profit restructuring, also received subpoenas. "They were going after folks to basically scare them into shutting up," Don Howard, who heads a charity called the James Irvine Foundation, said. (OpenAI claims that this was part of the standard legal process.)

Altman has long supported Democrats. "I'm very suspicious of powerful autocrats telling a story of fear to gang up on the weak," he told us. "That's a Jewish thing, not a gay thing." In 2016, he endorsed Hillary Clinton and called Trump "an unprecedented threat to America." In 2020, he donated to the Democratic Party and to the Biden Victory Fund. During the Biden Administration, Altman met with the White House at least half a dozen times. He helped develop a lengthy executive order laying out the first federal regime of safety tests and other guardrails for A.I. When Biden signed it, Altman called it a "good start."

In 2024, with Biden's poll numbers slipping, Altman's rhetoric began to shift. "I believe that America is going to be fine no matter what happens in this election," he said. After Trump won, Altman donated a million dollars to his inaugural fund, then took selfies with the influencers Jake and Logan Paul at the Inauguration. On X, in his standard lowercase style, Altman wrote, "watching @potus more carefully recently has really changed my perspective on him (i wish i had done more of my own thinking . . .)." Trump, on his first day back in office, repealed Biden's executive order on A.I. "He's found an effective way for the Trump Administration to do his bidding," a senior Biden Administration official said, of Altman.

Musk continues to excoriate Altman in public, calling him "Scam Altman" and "Swindly Sam." (When Altman complained on X about a Tesla he'd ordered, Musk replied, "You stole a non-profit.") And yet, in Washington, Altman seems to have outflanked him. Musk spent more than two hundred and fifty million dollars to help Trump get reëlected, and worked in the White House for months. Then Musk left Washington, damaging his relationship with Trump in the process.

Altman is now one of Trump's favored tycoons, even accompanying him on a trip to visit the British Royal Family at Windsor Castle. Altman and Trump speak a few times a year. "You can just, like, call him," Altman said. "This is not a buddy. But, yeah, if I need to talk to him about something, I will." When Trump hosted a dinner with tech leaders at the White House last year, Musk was notably absent; Altman sat across from the President. "Sam, you're a big leader," Trump said. "You told me things before that are absolutely unbelievable."

Over the years, Altman has continued to compare the quest for A.G.I. to the Manhattan Project. Like J. Robert Oppenheimer, who used impassioned appeals about saving the world from the Nazis to persuade physicists to uproot their lives and move to Los Alamos, Altman leverages fears about the geopolitical stakes of his technology. Depending on the audience, Altman has used this analogy to encourage either acceleration or caution. In a meeting with U.S. intelligence officials in the summer of 2017, he claimed that China had launched an "A.G.I. Manhattan Project," and that OpenAI needed billions of dollars of government funding to keep pace. When pressed for evidence, Altman said, "I've heard things." It was the first of several meetings in which he made the claim. After one of them, he told an intelligence official that he would follow up with

evidence. He never did. The official, after looking into the China project, concluded that there was no evidence that it existed: "It was just being used as a sales pitch." (Altman says that he does not recall describing Beijing's efforts in exactly that way.)

With more safety-conscious audiences, Altman invoked the analogy to imply the opposite: that A.G.I. had to be pursued carefully, with international coördination, lest the consequences be disastrous. In 2017, Amodei hired Page Hedley, a former public-interest lawyer, to be OpenAI's policy and ethics adviser. In an early PowerPoint presentation to executives, Hedley outlined how OpenAI might avert a "catastrophic" arms race—perhaps by building a coalition of A.I. labs that would eventually coördinate with an international body akin to NATO, to insure that the technology was deployed safely. As Hedley recalled it, Brockman didn't understand how this would help the company beat its competitors. "No matter what I said," Hedley told us, "Greg kept going back to 'So how do we raise more money? How do we win?' " According to several interviews and contemporaneous records, Brockman offered a counterproposal: OpenAI could enrich itself by playing world powers—including China and Russia—against one another, perhaps by starting a bidding war among them. According to Hedley, the thinking seemed to be, It worked for nuclear weapons, why not for A.I.?

He was aghast: "The premise, which they didn't dispute, was 'We're talking about potentially the most destructive technology ever invented—what if we sold it to Putin?' " (Brockman maintains that he never seriously entertained auctioning A.I. models to governments. "Ideas were batted around at a high level about what potential frameworks might look like to encourage cooperation between nations—something akin to an

International Space Station for AI," an OpenAI representative said. "Attempting to characterize it as anything more than that is utterly ridiculous.")

Brainstorming sessions often produce outlandish ideas. Hedley hoped that this one, which came to be known internally as the "countries plan," would be dropped. Instead, according to several people involved and to contemporaneous documents, OpenAI executives seemed to grow only more excited about it. Brockman's goal, according to Jack Clark, OpenAI's policy director at the time, was to "set up, basically, a prisoner's dilemma, where all of the nations need to give us funding," and that "implicitly makes not giving us funding kind of dangerous." A junior researcher recalled thinking, as the plan was detailed at a company meeting, "This is completely fucking insane."

Executives discussed the approach with at least one potential donor. But later that month, after several employees talked about quitting, the plan was abandoned. Altman "would lose staff," Hedley said. "I feel like that was always something that had more weight in Sam's calculations than 'This is not a good plan because it might cause a war between great powers.' "

"I cannot wait for crop tops to go out of style."

Cartoon by Meredith Southard

Open cartoon gallery

Undeterred by the collapse of the countries plan, Altman pursued variations on the theme. In January, 2018, he convened an "A.G.I. weekend" at the Hotel Bel-Air, an Old Hollywood resort with rolling gardens of pink bougainvillea and an artificial pond stocked with real swans. The attendees included Nick Bostrom, a philosopher, then at Oxford, who had become a prophet of A.I. doom; Omar Al Olama, an Emirati sultan and an A.I. booster; and at least seven billionaires. The safety-concerned among them were told that this would be an opportunity to think through how society might prepare for the disruptive arrival of artificial general intelligence; the investors arrived expecting to hear pitches.

The days were spent in a sleek conference room, where guests gave talks. (Hoffman, the LinkedIn co-founder, expounded on the possibilities of encoding A.I. with Buddhist compassion.) The final presenter was Altman, armed with a pitch deck that described a global cryptocurrency "redeemable for the attention of the AGI." Once the A.G.I. was maximally useful, and "anti-evil," people everywhere would clamor to buy time on OpenAI's servers. Amodei wrote in his notes, "This idea was absurd on its face (would Vladimir Putin end up owning some of the tokens? . . .) In retrospect this was one of many red flags about Sam that I should have taken more seriously." The plan seemed like a cash grab, but Altman sold it as a boon for A.I. safety. One of his slides read, "I want to get as many people on the 'good' team as possible, and win, and do the right thing." Another read, "Please hold your laughter until the end of the presentation."

Altman's fund-raising pitch has evolved over the years, but it has always reflected the fact that the development of A.G.I. requires a

staggering amount of capital. He was following a relatively simple "scaling law": the more data and computing power you used to train the models, the smarter they seemed to get. The specialized chips that enable this process are enormously expensive. OpenAI, in its most recent funding round alone, raised more than a hundred and twenty billion dollars—the largest private round in history, and a sum four times larger than the biggest I.P.O. ever. "When you think about entities with a hundred billion dollars they can discretionarily spend per year, there really are only a handful in the world," a tech executive and investor told us. "There's the U.S. government, and the four or five biggest U.S. tech companies, and the Saudis, and the Emiratis—that's basically it."

Altman's initial focus was Saudi Arabia. He first met Mohammed bin Salman, the country's crown prince and de-facto monarch, in 2016, at a dinner at San Francisco's Fairmont Hotel. After that, Hedley recalled, Altman referred to the prince as "a friend." In September, 2018, according to Hedley's notes, Altman said, "I'm trying to decide if we would ever take tens of billions from the Saudi PIF," or public investment fund.

The following month, a hit squad, reportedly acting on bin Salman's orders, strangled Jamal Khashoggi, a Washington *Post* journalist who had been critical of the regime, and used a bone saw to dismember his corpse. A week later, it was announced that Altman had joined the advisory board for Neom, a "city of the future" that bin Salman hoped to build in the desert. "Sam, you cannot be on this board," Clark, the policy director, who now works at Anthropic, recalled telling Altman. He initially defended his involvement, telling Clark that Jared Kushner had assured him that the Saudis "didn't do this." (Altman does not recall this. Kushner says that they were not in contact at the time.)

As bin Salman's role became increasingly clear, Altman left the Neom board. Yet behind the scenes, a policy consultant from whom Altman sought advice recalled, he treated the situation as a temporary setback, asking whether he could somehow still get money from bin Salman. "The question was not 'Is this a bad thing or not?' " the consultant said. "But, just, 'What would the consequences be if we did it? Would there be some export-control issue? Would there be sanctions? Like, can I get away with it?' "

By then, Altman was already eying another source of cash: the United Arab Emirates. The country was in the midst of a fifteen-year effort to transform itself from an oil state to a tech hub. The project was overseen by Sheikh Tahnoon bin Zayed al-Nahyan, the President's brother and the nation's spymaster. Tahnoon runs the state-controlled A.I. conglomerate G42, and controls $1.5 trillion in sovereign wealth. In June, 2023, Altman visited Abu Dhabi, meeting with Olama and other officials. In remarks at a government-backed function, he said that the country had "been talking about A.I. since before it was cool," and outlined a vision for the future of A.I. with the Middle East in a "central role."

Fund-raising from Gulf states has become customary for many large businesses. But Altman was pursuing a more sweeping geopolitical vision. In the fall of 2023, he began quietly recruiting new talent for a plan—eventually known as ChipCo—in which Gulf states would provide tens of billions of dollars for the construction of huge microchip foundries and data centers, some to be situated in the Middle East. Altman pitched Alexandr Wang, now the head of A.I. at Meta, on a leadership role, telling him that Jeff Bezos, the founder of Amazon, could head the new company. Altman sought enormous contributions from the Emiratis. "My

understanding was that this whole thing happened without any board knowledge," the board member said. A researcher Altman tried to recruit for the project, James Bradbury, recalled turning him down. "My initial reaction was 'This is gonna work, but I don't know if I want it to work,' " he said.

A.I. capacity may soon displace oil or enriched uranium as the resource that dictates the global balance of power. Altman has said that computing power is "the currency of the future." Normally, it might not matter where a data center was situated. But many American national-security officials were anxious about concentrating advanced A.I. infrastructure in Gulf autocracies. The U.A.E.'s telecommunications infrastructure is heavily dependent on hardware from Huawei, a Chinese tech giant linked to the government, and the U.A.E. has reportedly leaked American technology to Beijing in the past. Intelligence agencies worried that advanced U.S. microchips sent to the Emiratis could be used by Chinese engineers. Data centers in the Middle East are also more vulnerable to military strikes; in recent weeks, Iran has bombed American data centers in Bahrain and the U.A.E. And, hypothetically, a Gulf monarchy could commandeer an American-owned data center and use it to build disproportionately powerful models—a version of the "AGI dictatorship" scenario, but in an actual dictatorship.

After Altman's firing, the person he relied on most was Chesky, the Airbnb co-founder and one of Altman's fiercest loyalists. "Watching my friend stare into the abyss like that, it made me question some fundamental things about what it means to really run a company," Chesky told us. The following year, at a gathering of Y Combinator alumni, he gave an impromptu talk, which ended up lasting two hours. "It felt like a

group-therapy session," he said. The upshot was: Your instincts for how to run the company that you started are the best instincts, and anyone who tells you otherwise is gaslighting you. "You're not crazy, even though people who work for you tell you you are," Chesky said. Paul Graham, in a blog post about the speech, gave this defiant attitude a name: Founder Mode.

Since the Blip, Altman has been in Founder Mode. In February, 2024, the *Wall Street Journal* published a description of Altman's vision for ChipCo. He conceived of it as a joint entity funded by an investment of five to seven trillion dollars. ("fk it why not 8," he tweeted.) This was how many employees learned about the plan. "Everyone was, like, 'Wait, what?' " Leike recalled. Altman insisted at an internal meeting that safety teams had been "looped in." Leike sent a message urging him not to falsely suggest that the effort had been approved.

During the Biden Administration, Altman explored getting a security clearance to join classified A.I.-policy discussions. But staffers at the RANDCorporation, which helped coördinate the process, expressed concern. "He has been actively raising 'hundreds of billions of dollars' from foreign governments," one of them wrote. "The UAE recently gifted him a car. (I assume it was a very nice car.)" The staffer continued, "The only person I can think of who ever went thru the process with this magnitude of foreign financial ties is Jared Kushner, and the adjudicators recommended that he not be granted a clearance." Altman ultimately withdrew from the process. "He was pushing these transactional relationships, primarily with the Emiratis, that raised a lot of red flags for some of us," a senior Administration official involved in talks with Altman told us. "A lot of people in the Administration did not trust him a

hundred per cent."

When we asked Altman about gifts from Tahnoon, he said, "I'm not gonna say what gifts he has given me specifically. But he and other world leaders . . . have given me gifts." He added, "We have a standard policy, which applies to me as well, which is that every gift from any potential business partner is disclosed to the company." Altman has at least two hypercars: an all-white Koenigsegg Regera, worth about two million dollars, and a red McLaren F1, worth about twenty million dollars. In 2024, Altman was spotted driving the Regera through Napa. A few seconds of video made its way onto social media: Altman in a low-slung bucket seat, peering out the window of a gleaming white machine. A tech investor aligned with Musk posted the footage on X, writing, "I'm starting a nonprofit next."

In 2024, Altman took two OpenAI employees to visit Sheikh Tahnoon on his two-hundred-and-fifty-million-dollar superyacht, the Maryah. One of the largest such vessels in the world, the Maryah has a helipad, a night club, a movie theatre, and a beach club. Altman's employees apparently stood out amid Tahnoon's armed security detail, and at least one later told colleagues that he found the experience disconcerting. Altman, on X, later referred to Tahnoon as a "dear personal friend."

Cartoon by Juan Astasio

Altman continued to meet with the Biden Administration, which had enacted a policy requiring White House approval for the export of sensitive technology. Multiple Administration officials emerged from these meetings nervous about Altman's ambitions in the Middle East. He often made grandiose claims, according to those officials, including calling A.I. "the new electricity." In 2018, he said that OpenAI was planning to buy a fully functioning quantum computer from a company called Rigetti Computing. This was news even to other OpenAI executives in the room. Rigetti was not yet close to being able to sell a usable quantum computer. In a meeting, Altman claimed that by 2026 an extensive network of nuclear-fusion reactors across the United States would power the A.I. boom. The senior Administration official said, "We were, like, 'Well, that's, you know, news, if they made nuclear fusion work.' " The Biden Administration ultimately withheld approval. "We're not going to be building advanced chips in the U.A.E.," a leader at the Department of Commerce told Altman.

Four days before Trump's Inauguration, the *Wall Street Journal* reported, Tahnoon paid half a billion dollars to the Trump family in exchange for a stake in its cryptocurrency company. The following day, Altman held a twenty-five-minute call with Trump, during which they discussed announcing a version of a ChipCo, timed so that Trump could take credit for it. On Trump's second day in office, Altman stood in the Roosevelt Room and announced Stargate, a five-hundred-billion-dollar joint venture that aims to build a vast network of A.I. infrastructure across the U.S.

In May, the Administration rescinded Biden's export restrictions on A.I. technology. Altman and Trump travelled to the Saudi royal court to meet with bin Salman. Around the same time, the Saudis advertised the launch

of a giant state-backed A.I. firm in the kingdom, with billions to spend on international partnerships. About a week later, Altman laid out a plan for Stargate to expand into the U.A.E. The company plans to build a data-center campus in Abu Dhabi which is seven times larger than Central Park and consumes roughly as much electrical power as the city of Miami. "The truth of this is, we're building portals from which we're genuinely summoning aliens," a former OpenAI executive said. "The portals currently exist in the United States and China, and Sam has added one in the Middle East." He went on, "I think it's just, like, wildly important to get how scary that should be. It's the most reckless thing that has been done."

The erosion of safety commitments has become an industry norm. The founding premise of Anthropic was that, given the right structure and leadership, it could keep safety commitments from disintegrating under commercial pressure. One such commitment was a "responsible scaling policy," which obligated Anthropic to stop training more powerful models if it could not demonstrate that they were safe. In February, as the firm secured thirty billion dollars in new funding, it weakened that pledge. In some respects, Anthropic still emphasizes safety more than OpenAI does. But Clark, the former policy director, has said, "The system of capital markets says, Go faster." He added, "The world gets to make this decision, not companies." Last year, Amodei sent a memo to Anthropic employees, disclosing that the firm would seek investments from the United Arab Emirates and Qatar and acknowledging that this would likely enrich "dictators." (Like many authors, we are both parties in a class-action lawsuit alleging that Anthropic used our books without our permission to train its models. Condé Nast has opted into a settlement agreement with Anthropic

regarding the company's use of certain books published by Condé Nast and its subsidiaries.)

In 2024, Anthropic partnered with Palantir, one of Silicon Valley's most hawkish defense contractors, pushing its A.I. model, Claude, directly into the military ecosystem. Anthropic became the only A.I. contractor used in the Pentagon's most classified settings. Last year, the Pentagon awarded the company a further two-hundred-million-dollar contract. In January, the U.S. military launched a midnight raid that captured the Venezuelan President, Nicolás Maduro. According to the *Wall Street Journal*, Claude was used in the classified operation.

But tensions arose between Anthropic and the government. Years earlier, OpenAI had deleted from its policies a blanket ban on using its technology for "military and warfare." Eventually, Anthropic's rivals—including Google and xAI—agreed to provide their models to the military for "all lawful purposes." Anthropic, whose policies bar it from enabling fully autonomous weapons or domestic mass surveillance, resisted on these points, slowing negotiations for an overhauled deal. On a Tuesday in late February, Defense Secretary Pete Hegseth summoned Amodei to the Pentagon and delivered an ultimatum: the firm had until 5:01 P.M. that Friday to abandon those prohibitions. The day before the deadline, Amodei declined to do so. Hegseth tweeted that he would designate Anthropic a "supply-chain risk"—a devastating blacklist historically reserved for companies, like Huawei, that have ties to foreign adversaries—and made good on the threat days later.

Hundreds of employees at OpenAI and Google signed an open letter titled "We Will Not Be Divided," defending Anthropic. In an internal

memo, Altman wrote that the dispute was "an issue for the whole industry," and claimed that OpenAI shared Anthropic's ethical boundaries. But Altman had been in negotiations with the Pentagon for at least two days. Emil Michael, the Under-Secretary of Defense for Research and Engineering, had contacted Altman as he sought replacements for Anthropic. "I needed to hurry and find alternatives," Michael recalled. "I called Sam, and he was willing to jump. I think he's a patriot." Altman asked Michael, "What can I do for the country?" It appears that he already knew the answer. OpenAI lacked the security accreditation required for the classified systems in which Anthropic's technology was embedded. But a fifty-billion-dollar deal, announced that Friday morning, integrated OpenAI's technology into Amazon Web Services, a key part of the Pentagon's digital infrastructure. That night, Altman announced on X that the military would now be using OpenAI's models.

By some measures, Altman's maneuver has not hindered the company's success. The day he announced the deal, a new funding round increased OpenAI's value by a hundred and ten billion dollars. But many users deleted the ChatGPT app. At least two senior employees departed—one for Anthropic. At a staff meeting, Altman chastised employees who raised concerns. "So maybe you think the Iran strike was good and the Venezuela invasion was bad," he said. "You don't get to weigh in on that."

Several executives connected to OpenAI have expressed ongoing reservations about Altman's leadership and floated Fidji Simo, who was formerly the C.E.O. of Instacart and now serves as OpenAI's C.E.O. for AGI Deployment, as a successor. Simo herself has privately said that she believes Altman may eventually step down, a person briefed on a recent

discussion told us. (Simo disputes this. Instacart recently reached a settlement with the F.T.C., in which it admitted no wrongdoing but agreed to pay a sixty-million-dollar fine for alleged deceptive practices under Simo's leadership.)

Altman describes his shifting commitments as a by-product of his ability to adapt to changing circumstances—not a nefarious "long con," as Musk and others have alleged, but a gradual, good-faith evolution. "I think what some people want," he told us, is a leader who "is going to be absolutely sure of what they think and stick with it, and it's not going to change. And we are in a field, in an area, where things change extremely quickly." He defended some of his actions as the practice of "normal competitive business." Several investors we spoke to described Altman's detractors as naïve to expect anything else. "There is a group of fatalistic extremists that has taken the safety pill almost to a science-fiction level," Conway, the investor, told us. "His mission is measured by numbers. And, when you look at the success of OpenAI, it's hard to argue with the numbers."

But others in Silicon Valley think that Altman's behavior has created unacceptable managerial dysfunction. "It's more about a practical inability to govern the company," the board member said. And some still believe that the architects of A.I. should be evaluated more stringently than executives in other industries. The vast majority of people we spoke to agreed that the standards by which Altman now asks to be judged are not those he initially proposed. During one conversation, we asked Altman whether running an A.I. company came with "an elevated requirement of integrity." This was supposed to be an easy question. Until recently, when asked a version of it, his answer was a clear, unqualified yes. Now he

added, “I think there’s, like, a lot of businesses that have potential huge impact, good and bad, on society.” (Later, he sent an additional statement: “Yes, it demands a heightened level of integrity, and I feel the weight of the responsibility every day.”)

Of all the promises made at OpenAI’s founding, arguably the most central was its pledge to bring A.I. into existence safely. But such concerns are now often derided in Silicon Valley and in Washington. Last year, J. D. Vance, the former venture capitalist who is now the Vice-President, addressed a conference in Paris called the A.I. Action Summit. (It was previously called the A.I. Safety Summit.) “The A.I. future is not going to be won by hand-wringing about safety,” he said. At Davos this year, David Sacks, a venture capitalist who was serving as the White House’s A.I. and crypto czar, dismissed safety concerns as a “self-inflicted injury” that could cost America the A.I. race. Altman now calls Trump’s deregulatory approach “a very refreshing change.”

OpenAI has closed many of its safety-focussed teams. Around the time the superalignment team was dissolved, its leaders, Sutskever and Leike, resigned. (Sutskever co-founded a company called Safe Superintelligence.) On X, Leike wrote, “Safety culture and processes have taken a backseat to shiny products.” Soon afterward, the A.G.I.-readiness team, tasked with preparing society for the shock of advanced A.I., was also dissolved. When the company was asked on its most recent I.R.S. disclosure form to briefly describe its “most significant activities,” the concept of safety, present in its answers to such questions on previous forms, was not listed. (OpenAI said that its “mission did not change” and added, “We continue to invest in and evolve our work on safety, and will continue to make organizational changes.”) The Future of Life Institute, a think tank whose principles on

safety Altman once endorsed, grades each major A.I. company on "existential safety"; on the most recent report card, OpenAI got an F. In fairness, so did every other major company except for Anthropic, which got a D, and Google DeepMind, which got a D-.

"My vibes don't match a lot of the traditional A.I.-safety stuff," Altman said. He insisted that he continued to prioritize these matters, but when pressed for specifics he was vague: "We still will run safety projects, or at least safety-adjacent projects." When we asked to interview researchers at the company who were working on existential safety—the kinds of issues that could mean, as Altman once put it, "lights-out for all of us"—an OpenAI representative seemed confused. "What do you mean by 'existential safety'?" he replied. "That's not, like, a thing."

A.I. doomers have been pushed to the fringes, but some of their fears seem less fantastical with each passing month. In 2020, according to a U.N. report, an A.I. drone was used in the Libyan civil war to fire deadly munitions, possibly without oversight by a human operator. Since then, A.I. has only become more central to military operations around the world, including, reportedly, in the current U.S. campaign in Iran. In 2022, researchers at a pharmaceutical company tested whether a drug-discovery model could be used to find new toxins; within a few hours, it had suggested forty thousand deadly chemical-warfare agents. And many more mundane harms are already coming to pass. We increasingly rely on A.I. to help us write, think, and navigate the world, accelerating what experts call "human enfeeblement"; the ubiquity of A.I. "slop" makes life easier for scammers and harder for people who simply want to know what's real. A.I. "agents" are starting to act independently, with little or no human supervision. Days before the 2024 New Hampshire Democratic primary,

thousands of voters received robocalls from an A.I.-generated deepfake of Joe Biden's voice, telling them to save their votes for November and stay home—an act of voter suppression requiring virtually no technical expertise. OpenAI is now facing seven wrongful-death lawsuits, which allege that ChatGPT prompted several suicides and a murder. Chat logs in the murder case show that it encouraged a man's paranoid delusion that his eighty-three-year-old mother was surveilling and trying to poison him. Soon afterward, he fatally beat and strangled her and stabbed himself. (OpenAI is fighting the lawsuits, and says that it's continuing to improve its model's safeguards.)

As OpenAI prepares for its potential I.P.O., Altman has faced questions not only about the effect of A.I. on the economy—it could soon cause severe labor disruption, perhaps eliminating millions of jobs—but about the company's own finances. Eric Ries, an expert on startup governance, derided "circular deals" in the industry—for example, OpenAI's deals with Nvidia and other chip manufacturers—and said that in other eras some of the company's accounting practices would have been considered "borderline fraudulent." The board member told us, "The company levered up financially in a way that's risky and scary right now." (OpenAI disputes this.)

In February, we spoke again with Altman. He was wearing a drab-green sweater and jeans, and sat in front of a photograph of a NASA moon rover. He tucked one leg beneath him, then hung it over the arm of his chair. In the past, he said, his main flaw as a manager had been his eagerness to avoid conflict. "Now I'm very happy to fire people quickly," he had told us. "I'm happy to just say, 'We're gonna bet in this direction.' " Any employees who didn't like his choices needed "to leave."

He is more bullish than ever about the future. “My definition of winning is that people crazy uplevel—and the insane sci-fi future comes true for all of us,” he said. “I’m very ambitious as far as, like, my hope for humanity, and what I expect us all to achieve. I weirdly have very little personal ambition.” At times, he seemed to catch himself. “No one believes you’re doing this just because it’s interesting,” he said. “You’re doing it for power or for some other thing.”

Even people close to Altman find it difficult to know where his “hope for humanity” ends and his ambition begins. His greatest strength has always been his ability to convince disparate groups that what he wants and what they need are one and the same. He made use of a unique historical juncture, when the public was wary of tech-industry hype and most of the researchers capable of building A.G.I. were terrified of bringing it into existence. Altman responded with a move that no other pitchman had perfected: he used apocalyptic rhetoric to explain how A.G.I. could destroy us all—and why, therefore, he should be the one to build it. Maybe this was a premeditated masterstroke. Maybe he was fumbling for an advantage. Either way, it worked.

Not all the tendencies that make chatbots dangerous are glitches; some are by-products of how the systems are built. Large language models are trained, in part, on human feedback, and humans tend to prefer agreeable responses. Models often learn to flatter users, a tendency known as sycophancy, and will sometimes prioritize this over honesty. Models can also make things up, a tendency known as hallucination. Major A.I. labs have documented these problems, but they sometimes tolerate them. As models have grown more complex, some hallucinate with more persuasive fabrications. In 2023, shortly before his firing, Altman argued that

allowing for some falsehoods can, whatever the risks, confer advantages. "If you just do the naïve thing and say, 'Never say anything that you're not a hundred per cent sure about,' you can get a model to do that," he said. "But it won't have the magic that people like so much." ♦

Published in the print edition of the April 13, 2026, issue, with the headline "Moment of Truth."

New Yorker Favorites

- "The Giving Tree" is sadder than you remember.
- Sylvia Frumkin checked herself in to the Creedmoor Psychiatric Center. From there, her fate was out of her control.
- How Apollo 13 got lost on its way to the moon—then made it back.
- The dog that inherited twelve million dollars.
- An unrequited love story with J.F.K., Jr.
- Thebest books of 2026 so far.

Sign up for our daily newsletter to receive the best stories from *The New Yorker*.

Ronan Farrow, a contributing writer at The New Yorker, has won a Pulitzer Prize and a George Polk Award. His 2025 podcast, "Not a Very Good Murderer," is the basis for an HBO docuseries.

Andrew Marantz is a staff writer at The New Yorker and the author of "Antisocial: Online Extremists, Techno-Utopians, and the Hijacking of the American Conversation."

The New Yorker Classics Newsletter

Gems from the archive that deliver timeless pleasures with fresh insight

SIGN UP

By signing up, you agree to our user agreement (including class action waiver and arbitration provisions), and acknowledge our privacy policy.

READ MORE

ANNALS OF INQUIRY

The Pentagon Went to War with Anthropic. What's Really at Stake?

The Trump Administration wants Claude to act like an obedient soldier. But, if you ask for a killer robot, the company argues, you might get more than you bargained for.

By Gideon Lewis-Kraus

A REPORTER AT LARGE

He Helped Stop Iran from Getting the Bomb

A former C.I.A. officer says that he recruited scientists as part of the United States' effort to disrupt Iran's nuclear program.

By David D. Kirkpatrick

THE FINANCIAL PAGE

An Economist's Quest to Solve America's Wage Problem

Arindrajit Dube argues that the answer is empowering workers and setting mandatory wage standards across industries.

By John Cassidy

ANNALS OF WAR

What's Behind Trump's New World Disorder?

A foreign policy freed of liberal pretenses and imperial ambitions could lead to restraint—or, as the Iran attack shows, simply license hit-and-run belligerence.

By Daniel Immerwahr

LETTER FROM TRUMP'S WASHINGTON

Donald Trump Is Breaking Up with Europe

And the war in Iran is helping him do it.

By Susan B. Glasser

THE FINANCIAL PAGE

Who Struck It Rich in the Markets When Trump Postponed Bombing Iran?

A series of uncannily timed bets on the price of oil and stocks deserves a proper investigation. It's far from clear that they'll get one.

By John Cassidy

LETTER FROM TRUMP'S WASHINGTON

Donald Trump's Case for War Fails to Mention How to Win It

The President poses an existential question: Can everything be going according to the plan with Iran if there is no plan?

with Iran if there is no plan?

By Susan B. Glasser

Q. & A.

Why David Boies Thinks We Should Support Trump's Iran War

The prominent lawyer says that Democrats should get behind the President, and make sure that he finishes the job.

By Isaac Chotiner

Privacy Information